# Tighter Bounds in Triangular Systems

Sung Jae Jun,* Joris Pinkse† and Haiqing Xu‡

Center for Auctions, Procurements and Competition Policy

Department of Economics

The Pennsylvania State University

February 2009

We propose a new identification procedure for the nonparametric triangular model of Chesher (2005). The bounds on the conditional quantiles we obtain are necessarily at least as tight as Chesher's bounds, often tighter. In an example with a binary endogenous regressor, one of Chesher's bounds is trivial whereas we obtain point identification. There are other potential uses of the ideas put forth in this paper, including an extension of the Vytlacil and Yildiz (VY, 2007) results for binary endogenous regressors in a weakly separable triangular system. The extension allows for nonbinary discrete endogenous regressors and could improve on the accuracy of the VY estimator if the endogenous regressor is binary.

* (corresponding author) 608 Kern Graduate Building, University Park 16802, sjun@psu.edu

† joris@psu.edu; Joris Pinkse is an extramural fellow at Center, Tilburg University.

‡ hux100@psu.edu

## 1. Introduction

We propose a new identification procedure for the nonparametric triangular model of Chesher (2005). The bounds on the conditional quantiles we obtain are necessarily at least as tight as Chesher's bounds, often tighter. In an example with a binary endogenous regressor, one of Chesher's bounds is trivial whereas we obtain point identification. There are other potential uses of the ideas put forth in this paper, including an extension of the Vytlacil and Yildiz (VY, 2007) results for binary endogenous regressors in a weakly separable triangular system. The extension allows for nonbinary discrete endogenous regressors and could improve on the accuracy of the VY estimator if the endogenous regressor is binary.

In a triangular model, an outcome variable $\boldsymbol{y}$ ('earnings') depends on an endogenous regressor ('education') $\boldsymbol{x}$, an error term $\boldsymbol{u}$, and possibly some exogenous variables; we call this the *structural equation*. The endogenous regressor $\boldsymbol{x}$ in turn depends on (a vector of) instruments $\boldsymbol{z}$ ('demographics') and an error term $\boldsymbol{v}$ ('talent'); the *reduced form equation*. The form of both functional relationships is left unspecified, which makes the model nonparametric, albeit that certain monotonicity assumptions are made. Chesher's (2005) objective (and ours) is to obtain bounds for the conditional $u$–quantile $\psi$ of earnings for a given level of education $x$ and talent $v$.[1] The bounds for the marginal effect of changes in $x$ can be deduced from the bounds for $\psi$.

Chesher achieves his bounds by examining the conditional quantiles of $\boldsymbol{y}$ given $\boldsymbol{x} = x$ and $\boldsymbol{z} = z_j$, $j = 1, 2$, for some distinct $z_1, z_2$. He notes that conditioning on $\boldsymbol{x} = x$ and $\boldsymbol{z} = z_j$ is equivalent to conditioning on ($\boldsymbol{x} = x$ and) $\boldsymbol{v} \in V(x, z_j)$, where the set $V(x, z_j)$ is determined by the reduced form equation. If the values of $z_1$ and $z_2$ are such that all elements in $V(x, z_1)$ are less than $v$ and all elements of $V(x, z_2)$ are no less than $v$, then Chesher shows that conditioning on $V(x, z_1)$ and $V(x, z_2)$ yields respectively a lower and an upper bound to $\psi$. The existence of such values of $z_1$ and $z_2$ is a *rank condition* needed for Chesher's (2005) method. He suggests using all such pairs $\{(z_{1t}, z_{2t})\}$ in the support of $\boldsymbol{z}$ and using the largest lower bound and the smallest upper bound across $t$ as the ultimate bounds for $\psi$.

To illuminate the ideas developed in this paper, consider two pairs $(z_{11}, z_{21})$ and $(z_{12}, z_{22})$ satisfying Chesher's conditions. To produce a lower bound, Chesher would use the largest of the bounds generated by $V(x, z_{11})$ and $V(x, z_{12})$. We show that if $V(x, z_{11})$ is contained in $V(x, z_{12})$, then the conditional quantile of $\boldsymbol{y}$ given $\boldsymbol{x} = x$ and $\boldsymbol{v} \in V(x, z_{12}) - V(x, z_{11})$ is identified and may produce a tighter lower bound than conditioning on either $V(x, z_{11})$ or $V(x, z_{12})$; the same applies to the

[1] Chesher uses a different but equivalent definition of the object of interest.

upper bound.[2] Moreover, $V(x, z_{12}) - V(x, z_{11})$ can be entirely above $v$ even if both $V(x, z_{11})$ and $V(x, z_{12})$ contain elements less than $v$. Thus, our rank condition is weaker than Chesher's (2005). Finally, in addition to taking differences of $V$–sets one can take unions and these operations can be combined and iterated indefinitely to produce a class of $V$–sets that resembles a *Dynkin system* (Billingsley, 1995, p.41).

The consequence of this richer class of conditioning sets is that our bounds are always at least as tight and typically tighter than Chesher's. For binary $\boldsymbol{x}$ at least one of Chesher's (2005) bounds is trivial, i.e. the instruments do not provide any information. But given sufficient variation in $\boldsymbol{z}$, we show in an example that the new procedure allows for point identification.

We show in a separate section that the proposed procedure can accommodate vector–valued $\boldsymbol{x}$. The principles of the identification mechanism are effectively the same as in the scalar $\boldsymbol{x}$ case, albeit that there are notational and intuitional complications.

We establish our bounds formally, provide examples to explain its use and illustrate the procedure graphically.

Chesher (2005) formulates his identification method using a 'local' identification approach. Although his exposition is attractive, because our approach is more complex we use a global approach to maintain readability. It should be emphasized that our method can be formulated locally, also. In appendix C.1 we explain the differences between the local and global formulations.

The above–described idea has other applications. We discuss one such use in detail, namely that of VY's procedure. Like Chesher (2005), VY consider a nonparametric triangular system with a discrete endogenous regressor (specifically a binary regressor in VY's case). However, VY's model features a latent variable, a different monotonicity assumption, and VY's aim is to estimate a conditional mean instead of a conditional quantile. VYplace more emphasis on estimation than does Chesher. We do not discuss estimation in this paper.

Like Chesher, VY's approach uses conditioning sets. Using a procedure related to, but simpler than, the one described above for Chesher's model, we show that the class of conditioning sets one can use with VY's procedure can be enlarged substantially. For VY's model with a binary endogenous regressor, an implication is that a conditional mean estimator based on one of the newly generated sets could be more accurate than one based on VY's sets.

---

[2] In the binary $\boldsymbol{x}$ case, the difference invariably produces a tighter bound.

More important, however, is the fact that the proposed approach allows a VY–like estimation procedure to be used in cases that $\boldsymbol{x}$ is discrete but not binary. Although an extension to vector–valued $\boldsymbol{x}$ is in principle possible, we do not discuss such an extension here.

The implications of our approach in the VY–environment are not established formally, but explained in the text, using examples and graphical illustrations.

The paper is organized as follows. Section 2 establishes the main results of this paper for scalar–valued $\boldsymbol{x}$ in Chesher's (2005) model. These results are extended to vector–valued $\boldsymbol{x}$ in section 3. Section 4 contains a discussion of the use of the ideas propagated in this paper in the VY model.

## 2. Main Idea

We now explain our main idea for the scalar–valued endogenous regressor case. The general case is addressed later. Define the $\tau$–quantile of a random variable $\boldsymbol{y}$ as $Q_{\boldsymbol{y}}(\tau) = \inf\{y : \mathbb{P}[\boldsymbol{y} \leq y] \geq \tau\}$ and let conditional quantiles be similarly defined.

Consider the model

$$\begin{cases} \boldsymbol{y} & = g(\boldsymbol{x}, \boldsymbol{u}), \\ \boldsymbol{x} & = h(\boldsymbol{z}, \boldsymbol{v}), \end{cases} \tag{1}$$

where $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}$ are observables, $g, h$ are unknown functions and $\boldsymbol{u}, \boldsymbol{v}$ are errors. For now, $g, h, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{x}, \boldsymbol{y}$ are all scalar–valued; $\boldsymbol{z}$ has support $\mathscr{S}_z \subset \Re^{d_z}$. We refer to $\boldsymbol{x}$ as an endogenous regressor and $\boldsymbol{z}$ as a vector of instruments. Like in Chesher (2005) the purpose is to find bounds on $\psi(u, x, v) = Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(u|x, v)$, noting that $\boldsymbol{x}$ can be discrete.

We make the following assumptions.

**Assumption A.** $\boldsymbol{u}, \boldsymbol{v}$ *have standard uniform distributions.*

**Assumption B.** $g, h$ *are nondecreasing and left–continuous in their second argument for all values of their first argument.*

**Assumption C.** $\boldsymbol{u}, \boldsymbol{v}$ *are independent of* $\boldsymbol{z}$.

**Assumption D.** $\boldsymbol{u}$ *is positive regression dependent on* $\boldsymbol{v}$*, i.e. the conditional quantile of* $\boldsymbol{u}$ *given* $\boldsymbol{v}$*,* $Q_{\boldsymbol{u}|\boldsymbol{v}}(u|v)$ *is nondecreasing in* $v$ *for all values of* $u$.

Our set–up is largely the same as Chesher's (2005); we discuss the differences in appendix C.1. Assumption A and left–continuity (in assumption B) are normalizations. For instance, if $\boldsymbol{v}$ has a distribution different from a uniform then the definition of $h$ changes accordingly. Similarly, for any left–continuous $h$ there is an observationally equivalent right–continuous function.[3] Assumption D is a strong but essential assumption, as is weak monotonicity (assumption B). Assumption C is both strong and standard in this context. We have not yet made a *rank condition* at this point and delay its discussion until the description of our identification procedure. Please note that additional exogenous variables can be added to both $g$ and $h$ without affecting the results here; one can simply condition on each value of such shared exogenous variables before applying our procedure.

The procedure described below only makes sense considering values of $x, v$ for which $\psi$ is well–defined. What we mean by 'well–defined' is that there exists some value of $z \in \mathscr{S}_z$ for which $h(z, v) = x$. For instance, if $\boldsymbol{y}$ ='earnings,' $\boldsymbol{x}$ ='education,' (1=college, 0=no college) and $\boldsymbol{v}$='talent,' then $\psi(u, 1, 0)$ would be the $u$–quantile of earnings for people with a college education who are ex ante at the bottom of the talent pool. Since no such people exist, the relevance of $\psi(u, 1, 0)$ is questionable. In the current context, the best one can do for the lower bound is to take the bottom of the earnings distribution. For the upper bound to $\psi(u, 1, 0)$, find the smallest $v$–value for which $\exists z \in \mathscr{S}_z : h(z, v) = 1$ and use the upper bound of $\psi(u, 1, v)$ using the procedure described below. The procedure described below thus only applies to $x, v$ for which $\psi(u, x, v)$ is well–defined.

The discussion below presumes that $\boldsymbol{x}$ is discrete–valued, but none of the theoretical results depend on this presumption. We build up to our ultimate identification result in steps, including two examples.

**Lemma 1.** *For all* $u, v \in \mathscr{U} = (0, 1]$ *and* $x \in \mathscr{S}_x$, $\psi(u, x, v) = g\big(x, Q_{\boldsymbol{u}|\boldsymbol{v}}(u|v)\big)$.

*Proof.* See appendix A. □

We first introduce some notation, after which we explain the difference between Chesher's identification result and ours. For any set $V$ and any scalar $v$, $V \geq v$ means that no elements of $V$ are less than $v$. Let $Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(u|x, V)$ be the conditional $u$–quantile of $\boldsymbol{y}$ given that $\boldsymbol{x} = x$ and that $\boldsymbol{v} \in V$. Define $V(x, z) = \{v \in \mathscr{U} : h(z, v) = x\}$, which by the weak monotonicity and the left–continuity of $h$ in $v$, the independence of $\boldsymbol{v}$ and $\boldsymbol{z}$, and the fact that $\boldsymbol{v}$ has support $\mathscr{U}$, equals

$$V(x, z) = \big(P[\boldsymbol{x} < x|\boldsymbol{z} = z], P[\boldsymbol{x} \leq x|\boldsymbol{z} = z]\big].$$

Further, let $\mathscr{Z}_c^+(x, v) = \{z \in \mathscr{S}_z : V(x, z) \geq v\}$ and $\mathscr{Z}_c^-(x, v) = \{z \in \mathscr{S}_z : V(x, z) \leq v\}$.

[3] $h$ can be neither left–continuous nor right–continuous in $v$ only at countably many $v$–values, i.e. an irrelevant set of measure zero.

**Lemma 2.** *Suppose that* $\mathscr{Z}_c^-(x,v)$ *and* $\mathscr{Z}_c^+(x,v)$ *are nonempty. For all* $u \in \mathscr{U}$ *and all* $z^- \in \mathscr{Z}_c^-(x,v)$ *and* $z^+ \in \mathscr{Z}_c^+(x,v)$*,*

$$Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(u|x,z^-) \le \psi(u,x,v) \le Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(u|x,z^+). \tag{2}$$

*If* $\mathscr{Z}_c^-(x,v)$ *or* $\mathscr{Z}_c^+(x,v)$ *is empty, then the corresponding bound will be trivial.*

*Proof.* See appendix A. □

Lemma 2 is essentially Chesher's theorem 1. If one takes a supremum of the left most expression and an infimum of the right most expression in equation (2), then lemma 2 provides Chesher's minimum length interval; see theorem 3 and the discussion following it in Chesher (2005)). Note that the nonemptiness of $\mathscr{Z}_c^-(x,v)$ and $\mathscr{Z}_c^+(x,v)$ is equivalent to the existence of $z^-, z^+ \in \mathscr{S}_z$ such that

$$P[\boldsymbol{x} < x|\boldsymbol{z} = z^+] \ge v \quad \text{and} \quad P[\boldsymbol{x} \le x|\boldsymbol{z} = z^-] \le v, \tag{3}$$

which exactly corresponds to Chesher's (2005) rank condition. As pointed out by Chesher, for binary $\boldsymbol{x}$ at least one bound is necessarily trivial since $P[\boldsymbol{x} < 0|\boldsymbol{z} = z^+] = 0$ and $P[\boldsymbol{x} \le 1|\boldsymbol{z} = z^-] = 1$. However, we will show in the following that it is generally possible to obtain nontrivial bounds in the binary case and to tighten Chesher's bounds if they are nontrivial.

Consider $\mathscr{Z}^+(x,v) = \{z \in \mathscr{S}_z : \sup V(x,z) \ge v\}$ and $\mathscr{Z}^-(x,v) = \{z \in \mathscr{S}_z : \inf V(x,z) \le v\}$. Note that $\mathscr{Z}^+, \mathscr{Z}^-$ are much larger than $\mathscr{Z}_c^+, \mathscr{Z}_c^-$, respectively. Indeed, $\mathscr{Z}^+(x,v)$ consists of all $z$ for which the largest value in $V(x,z)$ is no less than $v$ whereas $\mathscr{Z}_c^+(x,v)$ requires *all* points in $V(x,z)$ to be no less than $v$.

For our bounds to be nontrivial we need $\mathscr{Z}^+(x,v)$ and $\mathscr{Z}^-(x,v)$ to be both nonempty, but it is not sufficient. What is sufficient but not necessary for the upper bound to be nontrivial is for there to exist $z_1^+, z_2^+ \in \mathscr{Z}^+(x,v)$ for which $V(x,z_1^+) \subset V(x,z_2^+)$ and $V(x,z_2^+) - V(x,z_1^+) \ge v$. The details of our rank condition will become apparent below.

Our result is based on the following two lemmas.

**Lemma 3.** *For all* $u \in \mathscr{U}$ *and all* $z \in \mathscr{S}_z$*,*

$$Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(u|x,z) = g\big\{x, Q_{\boldsymbol{u}|\boldsymbol{v}}\big(u|V(x,z)\big)\big\}. \tag{4}$$

*Proof.* See appendix A. □

**Lemma 4.** *For all* $u \in \mathscr{U}$ *and* $z_1, z_2 \in \mathscr{S}_z$*, if* $\mu$ *is the Lebesgue measure then*

(i) $V(x,z_1) \subset V(x,z_2),\ \mu\big(V(x,z_2) - V(x,z_1)\big) > 0 \Rightarrow g\big\{x, Q_{\boldsymbol{u}|\boldsymbol{v}}\big(u|V(x,z_2) - V(x,z_1)\big)\big\}$ *is identified,*

(ii) $V(x,z_1) \cap V(x,z_2) = \emptyset,\ \mu\big(V(x,z_1) \cup V(x,z_2)\big) > 0 \Rightarrow g\big\{x, Q_{\boldsymbol{u}|\boldsymbol{v}}\big(u|V(x,z_2) \cup V(x,z_1)\big)\big\}$ *is identified.*

*Proof.* The proof is in appendix A. □

The following example shows how lemma 4 can be valuable in practice.

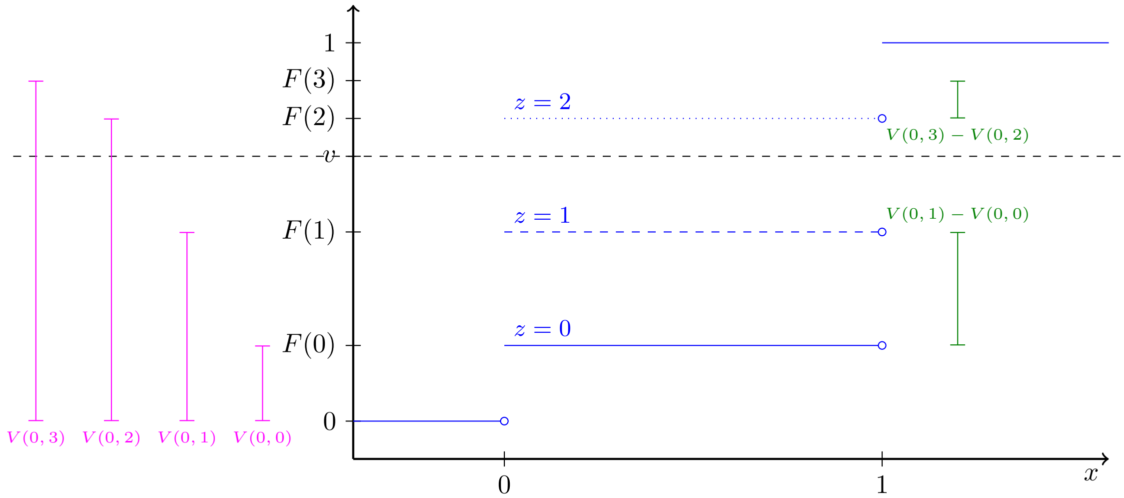


FIGURE 1. Example of how sets are combined.

**Example 1.** *Let $\mathscr{S}_x = \{0,1\}$ and suppose that $\boldsymbol{x} = I\big(\boldsymbol{v} \geq F(\boldsymbol{z})\big)$, where $F$ is some distribution function. Suppose first that $\mathscr{S}_z = \{0,1,2,3\}$ and concentrate on the case $x = 0$. Then $V(0,0) = \big(0, F(0)\big]$, $V(0,1) = \big(0, F(1)\big]$, $V(0,2) = \big(0, F(2)\big]$, and $V(0,3) = \big(0, F(3)\big]$. If $0 \leq F(0) < F(1) < v < F(2) < F(3) \leq 1$, then given the monotonicity assumptions the set $V(0,1) - V(0,0) = \big(F(0), F(1)\big]$ is likely to yield a tighter lower bound than any of the four $V(0,z)$–sets; see figure 1. Similarly, $V(0,3) - V(0,2)$ yields an upper bound, which none of the $V(0,z)$ sets can provide. The larger is the number of $z$–values, the tighter are the bounds, with point identifiYellowcation possible for continuous $\boldsymbol{z}$.* □

Lemma 4 is however not the best we can do. Indeed, the set differences and unions used in lemma 4 can be repeated, leading to a *Dynkin system* or $\lambda$ *system* (Billingsley, 1995, p.41) consisting solely of measurable sets.

**Definition 1.** *Let $\mathscr{A}$ be a collection of measurable subsets of $\mathscr{U}$. Then $\mathscr{D} = \mathscr{D}(\mathscr{A})$ is the collection $\mathscr{D}_\infty$ in the following iterative scheme. Let $\mathscr{D}_0 = \mathscr{A}$. Then for all $t \geq 0$, $\mathscr{D}_{t+1}$ consists of all sets $A^*$ such that at least one of the following three conditions is satisfied.*

(i) $A^* \in \mathscr{D}_t$,

(ii) $\exists A_1, A_2 \in \mathscr{D}_t : A_1 \subset A_2,\ \mu(A_2 - A_1) > 0,\ A^* = A_2 - A_1$,

(iii) $\exists A_1, A_2 \in \mathscr{D}_t : A_1 \cap A_2 = \emptyset,\ \mu(A_1 \cup A_2) > 0,\ A^* = A_1 \cup A_2$. □

Let $\mathscr{V}(x) = \bigcup_{z \in \mathscr{S}_z} \{V(x,z)\}$ and use $\mathscr{D}(x)$ as short hand for $\mathscr{D}\big(\mathscr{V}(x)\big)$. Let further $\mathscr{D}^-(x,v) = \big\{V \in \mathscr{D}(x) : V \le v\big\}$ and $\mathscr{D}^+(x,v) = \big\{V \in \mathscr{D}(x) : V \ge v\big\}$.

**Theorem 1.** *For all $u \in \mathscr{U}$,*

$$\sup_{V \in \mathscr{D}^-(x,v)} Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(u|x,V) \le \psi(u,x,v) \le \inf_{V \in \mathscr{D}^+(x,v)} Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(u|x,V). \tag{5}$$

*Proof.* See appendix B. □

For an illustration of the power of theorem 1, consider the following example.

**Example 2.** *Suppose that $h(z,v) = I\big(v \ge \Phi(z'\pi)\big)$, where $\Phi$ is the standard normal distribution function. Suppose that $\mathbb{E}[\boldsymbol{z}\boldsymbol{z}'] > 0$ and at least one element of the $\boldsymbol{z}$–vector has support $\Re$ and a nonzero $\pi$–coefficient. Then $\psi(u,x,v)$ is point–identified. Indeed, for $x = 0$, fix $u, v$ and note that $V(0,z) = \big(0, \Phi(z'\pi)\big]$. Let $\{z_t\}$ be a sequence such that $\forall t : \Phi(z_t'\pi) < \Phi(z_{t+1}'\pi)$ and $\lim_{t\to\infty} \Phi(z_t'\pi) = \Phi(z_\infty'\pi) = v$ for some $z_\infty \in \mathscr{S}_z$. Then $V_t = \big(\Phi(z_t'\pi), v\big] \in \mathscr{D}^-(0,v)$ and $\lim_{t\to\infty} Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(u|0,V_t) = \psi(u,0,v)$. Using the same procedure with $\forall t : \Phi(z_{t+1}'\pi) < \Phi(z_t'\pi)$, $\lim_{t\to\infty} \Phi(z_t'\pi) = v$, and $V_t = \big(v, \Phi(z_t'\pi)\big]$, the upper bound converges to $\lim_{t\to\infty} Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(u|0,V_t) = \lim_{\tilde{v}\downarrow v} \psi(u,0,\tilde{v})$.* □

Our rank condition, then, is that the collections $\mathscr{D}^-(x,v)$, $\mathscr{D}^+(x,v)$ are both nonempty. Our rank condition is sufficient for both bounds to be nontrivial. Moreover, even if Chesher's rank condition holds (and hence ours, also), our bounds are always at least equally tight and often tighter.

## 3. General Case

We again consider the model (1), albeit that $\boldsymbol{x}, \boldsymbol{v} \in \Re^{d_x}$ can now be vector–valued, $h(\boldsymbol{z},\boldsymbol{v})$ is a vector with $j$–th element $h_j(\boldsymbol{z}, \boldsymbol{v_j})$. The purpose is again to find bounds on $\psi(u,x,v)$ for values $x, v$ for which $\psi(u,x,v)$ is well–defined.

The next two assumptions replace assumptions A and B. We maintain assumptions C and D, albeit that their meaning has changed given that $\boldsymbol{v}, \boldsymbol{x}$ are now vector–valued.

**Assumption E.** *$\boldsymbol{u}, \boldsymbol{v_1}, \ldots, \boldsymbol{v_{d_x}}$ have standard uniform distributions.*

**Assumption F.** *$g$ is nondecreasing in $u$ for all values of $x$ and $h$ is for all $j = 1, \ldots, d_x$ nondecreasing in $v_j$ for all values of $u, v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_{d_x}$.*

For any set $V \subset \mathscr{U}^{d_x}$, $V \geq v$ means that for all values $\tilde{v} \in V$, $\tilde{v} \geq v$, i.e. $\tilde{v}_j \geq v_j$ for $j = 1, \ldots, d_x$. Using this notion of inequality for (sets of) vectors, generate $\mathscr{D}^-(x, v)$ and $\mathscr{D}^+(x, v)$ along the steps of section 2. Our rank condition in the multivariate case is the same as in the univariate case, albeit that $\mathscr{D}^-(x, v)$ and $\mathscr{D}^+(x, v)$ now contain multidimensional $V$–sets.

**Theorem 2.** *For all $u \in \mathscr{U}$,*

$$\sup_{V \in \mathscr{D}^-(x,v)} Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(u|x, V) \leq \psi(u, x, v) \leq \inf_{V \in \mathscr{D}^+(x,v)} Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}}(u|x, V). \tag{6}$$

*Proof.* See appendix B. □

Although the theory is largely the same as when $\boldsymbol{x}$ is scalar–valued, the construction of the $\mathscr{D}^-(x, v)$–collections is intuitively more complicated. We now use a simple example to illustrate our method of obtaining bounds.

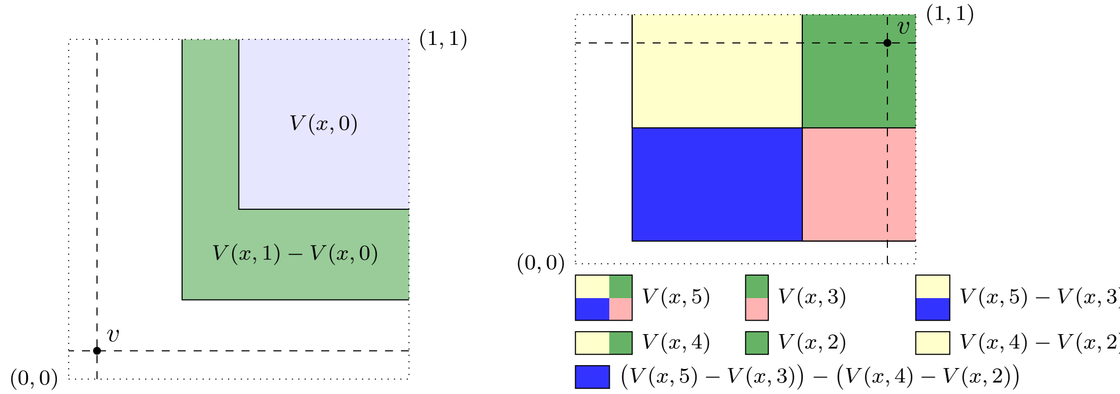


Figure 2. How to obtain upper and lower bounds when $x = (1, 1)$.

**Example 3.** *Let $\boldsymbol{x} \in \Re^2$ consist of binary random variables. Pick some value $v$, and consider the problem of getting an upper bound for $\psi(u, x, v)$ for $x = (1, 1)$. This case is illustrated in the left graph of figure 2. With Chesher's (2005) procedure, using $z = 0, 1$ one would (at best) be conditioning on the set of $v$'s consisting of the two shaded regions combined, i.e. on individuals who have talent in two dimensions exceeding a certain lower bound. Our procedure, in contrast, would be using as an upper bound the L–shaped set of $v$–values $V(x, 1) - V(x, 0)$, which yields a tighter upper bound by the monotonicity assumption.*

*Obtaining a lower bound here requires an area that is entirely to the left and below $v$, for which at least four $z$–values are needed. This can be accomplished by using the four sets of $v$'s $V(x, 2)$, $V(x, 3)$,*

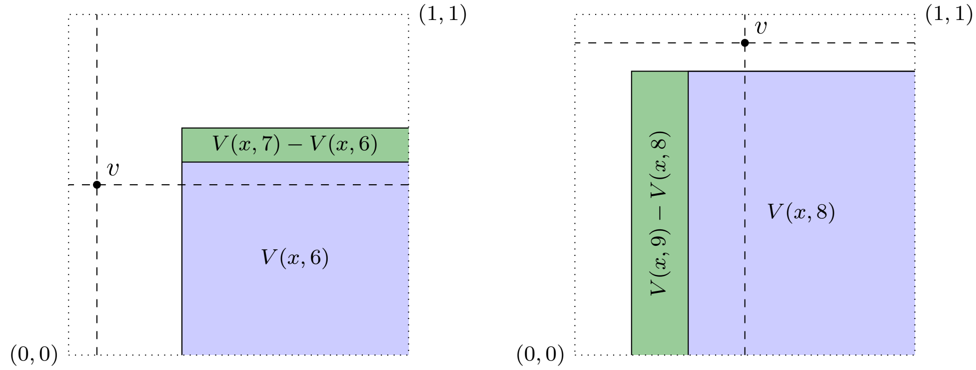


FIGURE 3. How to obtain upper and lower bounds when $x = (1, 0)$.

*$V(x,4)$, $V(x,5)$ indicated in the right graph of figure 2. Again, $V(x,5)-V(x,4)-V(x,3)+V(x,2)$ is a more favorable set to condition on since with Chesher's identification procedure only the trivial lower bound can be obtained.*

*The situation for $x = (1,0)$ is illustrated in figure 3. Chesher's procedure yields trivial lower and upper bounds, whereas ours yields nontrivial upper and lower bounds by conditioning on $V(x,7)-V(x,6)$ and $V(x,9)-V(x,8)$, respectively.* □

## 4. Another Use

The principle underlying the identification methodology developed above has other uses. An example is the identification strategy employed by VY. We show here that the class of conditioning sets used in VY can be broadened, which can lead to efficiency improvements in their binary endogenous regressor case and also allows their method to be extended to the case of nonbinary discrete endogenous regressors.

To maintain notational consistency, we express the VY model as

$$\begin{cases} \boldsymbol{y} &= g(\boldsymbol{\ell}, \boldsymbol{u}), \\ \boldsymbol{\ell} &= m(\boldsymbol{x}, \boldsymbol{w}), \\ \boldsymbol{x} &= h(\boldsymbol{z}, \boldsymbol{v}), \end{cases} \tag{7}$$

where $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\ell}$ are unobserved, $\boldsymbol{v}$ has a uniform $(0,1]$–distribution, $\boldsymbol{u}, \boldsymbol{v}$ are independent of $\boldsymbol{w}, \boldsymbol{z}$, $h$ is left–continuous and nondecreasing in its second argument and $\boldsymbol{x}$ is discrete. Further, $\theta(\ell, v) = \mathbb{E}[g(\ell, \boldsymbol{u})|\boldsymbol{v} = v]$ is strictly increasing in $\ell$ for all $v$.

Let $\boldsymbol{y}(x) = g\big(m(x, \boldsymbol{w}), \boldsymbol{u}\big)$. Because $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x})$, $\boldsymbol{y}(x)$ is a counterfactual if $\boldsymbol{x} \neq x$. The objective is to identify (and in VY to estimate)

$$\delta(x, w) = \mathbb{E}[\boldsymbol{y}(x)|\boldsymbol{w} = w],$$

from which one could e.g. infer treatment effects. In VY, $\boldsymbol{x}$ is binary and $h$ is taken to be $I\big(\boldsymbol{v} > \eta(\boldsymbol{z})\big)$ for $\eta(z) = \mathbb{P}[\boldsymbol{x} = 0|\boldsymbol{z} = z]$. The methodology, as described in VY, only applies to the binary $\boldsymbol{x}$ case.

We now return to the more general model (7) and show how identification obtains. Note first that

$$\mathbb{E}[\boldsymbol{y}|\boldsymbol{x} = x, \boldsymbol{w} = w, \boldsymbol{z} = z] = \mathbb{E}[\boldsymbol{y}(x)|\boldsymbol{w} = w, \boldsymbol{v} \in V(x, z)] = \mathbb{E}\big[g\big(m(x, w), \boldsymbol{u}\big)|\boldsymbol{v} \in V(x, z)\big],$$

where $V(x, z)$ is as defined before. Consequently, using arguments similar to those in lemma 4,

$$\mathscr{E}(x, w, V) = \mathbb{E}[\boldsymbol{y}(x)|\boldsymbol{w} = w, \boldsymbol{v} \in V] = \mathbb{E}\big[g\big(m(x, w), \boldsymbol{u}\big)|\boldsymbol{v} \in V\big] \tag{8}$$

is for all $V \in \mathscr{D}(x)$ (see section 2) identified. Now, if $V^c = \mathscr{U} - V$ then

$$\delta(x, w) = \mathscr{E}(x, w, V)\mu(V) + \mathscr{E}(x, w, V^c)\mu(V^c). \tag{9}$$

As noted by VY, in VY's binary model, monotonicity of $\theta$ is not needed for identification. Indeed, if $\inf_{z \in \mathscr{S}_z} \eta(z) = 0$ then one can construct a sequence $z_1, z_2, \ldots$ such that $\lim_{t\to\infty} \eta(z_t) = 0$, which implies that $\lim_{t\to\infty} V(1, z_t) = \mathscr{U}$ and hence $\delta(1, w)$ is then identified. Implementing such an *identification at infinity* (see also Heckman, 1990) argument in practice would mean that (still aside from the condition on $\eta$) one would use only an asymptotically negligble fraction of the data, which is inefficient.

We now proceed to discuss the case in which $\mathscr{D}(x)$ contains no set with Lebesgue measure one or in which greater efficiency is desired, making use of VY's monotonicity assumption. Suppose that $V \in \mathscr{D}(x)$, that $V^c \in \mathscr{D}(x^*)$ for some $x^* \neq x$, and that there exists some $w^*$ for which $m(x, w) = m(x^*, w^*)$. Then

$$\mathscr{E}(x, w, V^c) = \mathbb{E}\big[g\big(m(x, w), \boldsymbol{u}\big)|\boldsymbol{v} \in V^c\big] = \mathbb{E}\big[g\big(m(x^*, w^*), \boldsymbol{u}\big)|\boldsymbol{v} \in V^c\big] = \mathscr{E}(x^*, w^*, V^c), \tag{10}$$

which is identified since $V^c \in \mathscr{D}(x^*)$.

The only question remaining is how to find such $w^*$. We again use an extension of the VY idea; a similar approach in a different context is in Pinkse (2001). Suppose that a set $\check{V}$ exists for which $\mu(\check{V}) > 0$ and $\check{V} \in \mathscr{D}(x) \cap \mathscr{D}(x^*)$. Then by (8) we can identify for all $w, w^*$

$$\mathbb{E}\big[\big\{g\big(m(x, w), \boldsymbol{u}\big) - g\big(m(x^*, w^*), \boldsymbol{u}\big)\big\}I(\boldsymbol{v} \in \check{V})\big], \tag{11}$$

which is zero if and only if $m(x,w) = m(x^*,w^*)$ by the strict monotonicity of $\theta$.

The presentation here is different and more general than the one in VY. For binary $\boldsymbol{x}$, VY only allow contiguous $V^*$–sets (see appendix C.2). As mentioned before, VY's results do not cover nonbinary $\boldsymbol{x}$.

The above discussion presumes that there is a single $x^*$–value such that $V^c \in \mathscr{D}(x^*)$. This is not necessary. Indeed, the same argument can be naturally extended to the case in which there are distinct values $x_1^*, \ldots, x_T^*$ different from $x$ and disjoint sets $V_t^* \in \mathscr{D}(x_t^*)$, $t = 1, \ldots, T$, such that $V^c = \bigcup_{t=1}^T V_t^*$.

Indeed, replace (9) with

$$\delta(x,w) = \mathscr{E}(x,w,V)\mu(V) + \sum_{t=1}^{T} \mathscr{E}(x,w,V_t^*)\mu(V_t^*). \tag{12}$$

We now explain how to obtain $\mathscr{E}(x,w,V_t^*)$ for each $t = 1, \ldots, T$. Use (10) with $V_t^*$ in lieu of $V^c$. We are left to find $w_t^*$. Let $\breve{V}_t \in \mathscr{D}(x) \cap \mathscr{D}(x_t^*)$ with $\mu(\breve{V}_t) > 0$. Apply (11) with $\breve{V}_t$ replacing $\breve{V}$.

The above discussion assumes the availability of $(x_t^*, V_t^*, \breve{V}_t)$–combinations which satisfy the necessary conditions. In the binary case (VY), it is sufficient to have $\eta$ vary with $z$ for this assumption to hold since for $\eta(z_1) < \eta(z_2)$, $\bigl(\eta(z_1), \eta(z_2)\bigr]$ belongs to both $\mathscr{D}(0)$ and $\mathscr{D}(1)$.[4]

We conclude with an example in which our approach is implemented in a case with nonbinary $\boldsymbol{x}$.

**Example 4.** *Let $h(z,v) = I\bigl(v > \eta_1(z)\bigr) + I\bigl(v > \eta_2(z)\bigr)$ for two functions $\eta_1, \eta_2$ such that for all $z \in \mathscr{S}_z$, $0 = \eta_0(z) < \eta_1(z) < \eta_2(z) < \eta_3(z) = 1$. If $\eta_1(z) = \Phi(z'\pi)$ and $\eta_2(z) = \Phi(z'\pi + \tilde{\pi})$ for some $\pi, \tilde{\pi}$ then one has an ordered probit model; we do not assume this.*

*Note that $V(x,z) = \bigl(\eta_x(z), \eta_{x+1}(z)\bigr]$. If $\eta_1, \eta_2$ vary continuously with $z$ and $\boldsymbol{z}$ is continuous,*[5] *then to identify $\delta(x,w)$ for $x = 0,1,2$ it is sufficient for $\bar{v}_1 > \underline{v}_2$, where $\bar{v}_x = \sup_z \eta_x(z)$ and $\underline{v}_x = \inf_z \eta_x(z)$.*[6] *The condition $\bar{v}_1 > \underline{v}_2$ means that there should be sufficient variation in $\eta_1, \eta_2$, i.e. there must be $z_1, z_2 \in \mathscr{S}_z$ for which $\eta_2(z_1) < \eta_1(z_2)$, which is essentially a rank condition.*

*For $x = 0$ (and likewise $x = 2$), see figure 4. Note that $\mathscr{D}(0)$ contains the set $V = (0, v^*]$ for $v^* > \underline{v}_2$, such that $V^c = (v^*, 1] \in \mathscr{D}(2)$. To find $w^*$ for given $w$, note that for $\underline{v}_2 < \underline{v}^* < \bar{v}^* < \bar{v}_1$, we can take $\breve{V} = (\underline{v}^*, \bar{v}^*] \in \mathscr{D}(0) \cap \mathscr{D}(2)$; in figure 4 we have taken $\underline{v}^*$ slightly greater than $\underline{v}_2$ and $\bar{v}^*$ slightly less than $\bar{v}_1$. Please note that $\breve{V}$ need not be taken as an interval. The set $\breve{\breve{V}}$ in figure 4 could also be used.*

[4] It belongs to $\mathscr{D}(0)$ because $\bigl(0, \eta(z_1)\bigr]$ and $\bigl(0, \eta(z_2)\bigr]$ belong to $\mathscr{D}(0)$ and to $\mathscr{D}(1)$ because $\bigl(\eta(z_1), 1\bigr]$ and $\bigl(\eta(z_2), 1\bigr]$ belong to $\mathscr{D}(1)$.

[5] Continuity of the entire $\boldsymbol{z}$–vector is not necessary, but it simplifies the discussion.

[6] We implicitly assume here that the support of $\boldsymbol{z}$ does not depend on $\boldsymbol{w}$, which can be relaxed at the expense of obfuscating notation.

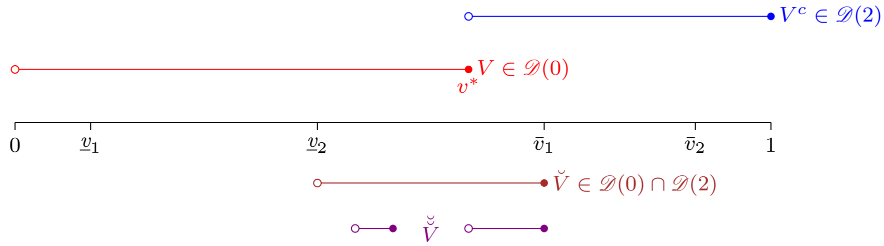


FIGURE 4. Ordered response example; $x = 0$.

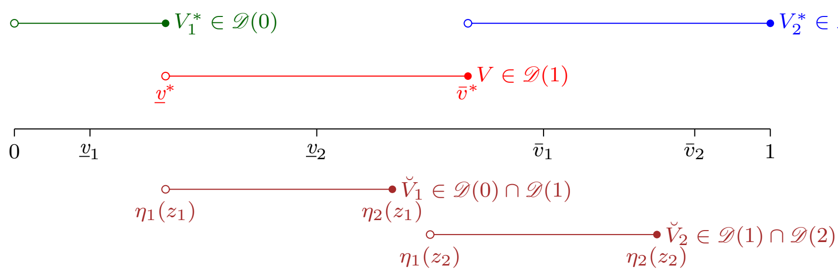


FIGURE 5. Ordered response example; $x = 1$.

*For $x = 1$ (see figure 5), $\mathscr{D}(1)$ contains the set $V = (\underline{v}^*, \bar{v}^*]$ where $\underline{v}^* < \bar{v}^*$ are such that $\underline{v}^* < \bar{v}_1$ and $\bar{v}^* > \underline{v}_2$. Then take $V_1^* = (0, \underline{v}^*] \in \mathscr{D}(0)$ and $V_2^* = (\bar{v}^*, 1] \in \mathscr{D}(2)$, such that $V_1^* \cup V_2^* = V^c$. A possible choice for $(\breve{V}_1, \breve{V}_2)$ is indicated in figure 5. Note that nonoverlapping $\breve{V}_1, \breve{V}_2$ necessarily exist by the rank condition; overlapping $\breve{V}_1, \breve{V}_2$ usually exist, also.*

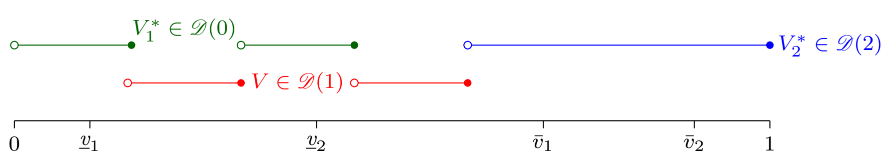


FIGURE 6. Ordered response example; $x = 1$, noncontiguous $V$.

*Again, noncontiguous $\breve{V}_1, \breve{V}_2$ are possible. For the purpose of estimation, it is likely best to choose $\breve{V}_1, \breve{V}_2$ as large as possible. This is not generally true for $V$ since choosing a larger $V$ means that $V_1^*, V_2^*$ will (together) be smaller. Indeed, there is no reason to believe that the optimal choice of $V$ for estimation purposes is a contiguous set. A noncontiguous choice of $V$ is shown in figure 6. The choice of $\breve{V}_1, \breve{V}_2$ is not impacted by the choice of $V, V_1^*, V_2^*$.* □

## References Cited


Billingsley, Patrick (1995), Probability and measure, third edition, Wiley, New York.

Chesher, Andrew (2005), Nonparametric identification under discrete variation, *Econometrica* 73, 1525–1550.

Heckman, James (1990), Selectivity bias: new developments, *American Economic Review* 80–2, 313–318.

Pinkse, Joris (2001), Nonparametric regression estimation using weak separability, UBC mimeo.

Vytlacil, Edward and Neșe Yildiz (2007), Dummy endogenous variables in weakly separable models, *Econometrica* 75, 757–779.


## Appendix A. Proofs of lemmas

*Proof of Lemma 1.* By the weak monotonicity assumption on $g$, $\psi(u,x,v) = g\big(x, Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{v}}(u|x,v)\big)$. If $\mathscr{Z}(x,v) = \{z : h(z,v) = x\}$ then $\boldsymbol{x} = x, \boldsymbol{v} = v \Leftrightarrow \boldsymbol{z} \in \mathscr{Z}(x,v), \boldsymbol{v} = v$, such that $Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{v}}(u|x,v) = Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}\big(u|v,\mathscr{Z}(x,v)\big)$, which equals $Q_{\boldsymbol{u}|\boldsymbol{v}}(u|v)$ by the independence of $(\boldsymbol{u},\boldsymbol{v})$ and $\boldsymbol{z}$. □

*Proof of Lemma 2.* We establish the upper bound; the argument for the lower bound is virtually identical. Note that $V(x,z^+) \geq v$ and that by positive regression dependence hence $Q_{\boldsymbol{u}|\boldsymbol{v}}(u|v) \leq Q_{\boldsymbol{u}|\boldsymbol{v}}\big(u|V(x,z^+)\big)$, which by independence of errors and instruments equals $Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}\big(u|V(x,z^+),z^+\big) = Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{z}}(u|x,z^+)$. Using lemma 1, we then have

$$\psi(u,x,v) = g\big(x, Q_{\boldsymbol{u}|\boldsymbol{v}}(u|v)\big) \leq g\big(x, Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{z}}(u|x,z^+)\big) = Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(u|x,z^+).$$ □

*Proof of Lemma 3.* By the weak monotonicity of $g$, $Q_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}}(u|x,z) = g\big(x, Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{z}}(u|x,z)\big)$. Since $\boldsymbol{x} = x, \boldsymbol{z} = z \Leftrightarrow \boldsymbol{v} \in V(x,z), \boldsymbol{z} = z$ and by the independence of errors and instruments,

$$Q_{\boldsymbol{u}|\boldsymbol{x},\boldsymbol{z}}(u|x,z) = Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}\big(u|V(x,z),z\big) = Q_{\boldsymbol{u}|\boldsymbol{v}}\big(u|V(x,z)\big).$$ □

*Proof of Lemma 4.* We show (i); (ii) is analogous. Note first that for all $y \in \Re$,

$$P[\boldsymbol{y} \leq y|\boldsymbol{x} = x, \boldsymbol{z} = z] = P[g(x,\boldsymbol{u}) \leq y|\boldsymbol{v} \in V(x,z)],$$

such that if $\mu^*(x,z) = \mu\big(V(x,z)\big)$,

$$P[\boldsymbol{y} \leq y|\boldsymbol{x} = x, \boldsymbol{z} = z_1^+]\mu^*(x,z_1^+) + P[g(x,\boldsymbol{u}) \leq y|\boldsymbol{v} \in V(x,z_2^+) - V(x,z_1^+)]\big(\mu^*(x,z_2^+) - \mu^*(x,z_1^+)\big)$$
$$= P[\boldsymbol{y} \leq y|\boldsymbol{x} = x, \boldsymbol{z} = z_2^+]\mu^*(x,z_2^+). \quad (13)$$

Since all quantiles of $\boldsymbol{y}$ given $\boldsymbol{x} = x, \boldsymbol{z} = z$ are identified and $\boldsymbol{v}$ is uniformly distributed, both the first left hand side term and the right hand side in (13) are identified, and hence so is the second left hand side term. Because the second left hand side term is identified for all values of $y$, so are the quantiles of $g(x, \boldsymbol{u})$ given $\boldsymbol{v} \in V(x, z_2^+) - V(x, z_1^+)$. $\square$

## Appendix B. Proofs of theorems

*Proof of Theorem 1.* Repeat the subtraction and union operations as per lemma 4. $\square$

## Appendix C. Miscellaneous

C.1. **Chesher (2005).** Chesher's (2005) conditions are expressed locally. In our notation Chesher starts from the model

$$\boldsymbol{y} = g(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}), \tag{14}$$

$$\boldsymbol{x} = h(\boldsymbol{z}, \boldsymbol{v}). \tag{15}$$

In restriction A1 of section 2.1, Chesher (2005) assumes that $g$ and $h$ are nondecreasing and left–continuous in $\boldsymbol{u}$ and $\boldsymbol{v}$. He then sets the function $h$ equal to the conditional $v$–quantile of $\boldsymbol{x}$ given $\boldsymbol{z}$, which is equivalent to our conditions on $h$, independence of $\boldsymbol{z}$ and $\boldsymbol{v}$ and $\boldsymbol{v}$ being standard uniform. Although restriction A1 does not require that $\boldsymbol{u}$ be standard uniform, this difference is meaningless, because a monotone transformation of $\boldsymbol{u}$ can be taken without loss of generality.

So the most apparent discrepancies between Chesher's (2005) conditions and ours are (i) the fact that $\boldsymbol{z}$ is in the function $g$, (ii) that $\boldsymbol{u}$ is not assumed independent of $\boldsymbol{z}$, and (iii) the positive regression dependence assumption. Before we address these differences, note first that in section 2.1, Chesher (2005) only considers quantiles at two different values of $\boldsymbol{z}$, say 0 and 1. Since other values of $\boldsymbol{z}$ are irrelvant to Chesher's analysis in section 2.1, we will treat $\boldsymbol{z}$ as binary for now.

Chesher's (2005) conditions are stated in restrictions B1–B4. Restriction B1 is a rank condition, which is discussed in detail in the main text of this paper. Restriction B2 corresponds to our condition of (positive) regression dependence. Although restriction B2 of Chesher (2005) does not require regression dependence at values of $v$ which are irrelevant for the construction of the bounds, this difference is not essential; we could (but do not) impose a similar condition at the expense of considerably more complicated notation.

Restrictions B3–B4 are invariance conditions of $Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}(u|v, z)$ and $g(x, z, u)$ with respect to $z$. These restrictions are local versions of our independence conditions. For instance, restriction B3 of Chesher requires that $Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}(u|v, 0) = Q_{\boldsymbol{u}|\boldsymbol{v},\boldsymbol{z}}(u|v, 1)$ for the values of interest $u, v$ only, which in essence assumes that $\boldsymbol{u}$ is 'locally independent' of $\boldsymbol{z}$ at $u$ given $\boldsymbol{v} = v$. We need the same condition, but assume full independence of $\boldsymbol{z}$ and $(\boldsymbol{v}, \boldsymbol{v})$ for the sake of parsimony. The distiction is of relevance if a single $u$–conditional quantile is desired instead of the full conditional quantile function (with argument $u$). Likewise, restriction B4 of Chesher requires that $g(x,1,u) = g(x,0,u)$ only at $x,u$ as opposed to at all $x,u$, which implies that for such $(x,u)$ $\boldsymbol{z}$ can be dropped as an argument of $g$.

Finally, in section 2.3 Chesher shows that there exists a data–generating process satisfying his model conditions such that his identified interval consists of a singleton. However, this result does not imply that his bounds are the tightest ones for a given data–denerating process. In fact, the main difference with Chesher's work arises at this point. In section 2.4 Chesher looks at bounds that can be obtained from the use of all pairs of $\boldsymbol{z}$–values, but we allow for the use of combinations of more than two $\boldsymbol{z}$–values with corresponding implications for the assumptions.

C.2. **Vytlacil and Yildiz (2007).** The method used in VY to find $w^*$ is as follows. Take $x = 0$ and $x^* = 1$ (the reverse case is symmetric). VY use two values of $\boldsymbol{z}$, $z_1, z_2$, such that $\eta(z_2) > \eta(z_1)$. For $V_1 = V(0, z_1) = \big(0, \eta(z_1)\big]$ and $V_2 = V(0, z_2) = \big(0, \eta(z_2)\big]$, we have $V_1^c = V(1, z_1) = \big(\eta(z_1), 1\big]$ and $V_2^c = V(1, z_2) = \big(\eta(z_2), 1\big]$, such that we can identify (and estimate)

$$
\begin{aligned}
\mathscr{E}(0, w, V_1)\mu(V_1) - \mathscr{E}(0, w, V_2)\mu(V_2) - \mathscr{E}(1, w^*, V_1^c)\mu(V_1^c) + \mathscr{E}(1, w^*, V_2^c)\mu(V_2^c)& \\
= \mathbb{E}\big[g\big(m(0,w), \boldsymbol{u}\big) I(\boldsymbol{v} \in V_1)\big] - \mathbb{E}\big[g\big(m(0,w), \boldsymbol{u}\big) I(\boldsymbol{v} \in V_2)\big]& \\
+ \mathbb{E}\big[g\big(m(1,w^*), \boldsymbol{u}\big) I(\boldsymbol{v} \in V_1^c)\big] - \mathbb{E}\big[g\big(m(1,w^*), \boldsymbol{u}\big) I(\boldsymbol{v} \in V_2^c)\big]& \\
= \mathbb{E}\big[\big\{g\big(m(0,w), \boldsymbol{u}\big) - g\big(m(1,w^*), \boldsymbol{u}\big)\big\} I(\boldsymbol{v} \in V_1 - V_2)\big],& \quad (16)
\end{aligned}
$$

which by the strict monotonicity of $\theta$ is zero if and only if $m(0, w) = m(1, w^*)$.

With our approach we note that $V_1 - V_2 \in \mathscr{D}(0) \cap \mathscr{D}(1)$ and apply (11). VY's approach is limited to sets $V_1, V_2$ for which $z_1, z_2$ exist such that $V_1 = V(x, z_1), V_2 = V(x, z_2)$ and $V_1^c = V(x^*, z_1), V_2^c = V(x^*, z_2)$. We allow for all sets $V_1, V_2$ for which $V_1 - V_2 \in \mathscr{D}(0) \cap \mathscr{D}(1)$ and $\mu(V_1 - V_2) > 0$, which is a collection of sets that is orders of magnitude larger.